

## Estimating the effects of excess nutrients on stream invertebrates from observational data

LESTER L. YUAN<sup>1</sup>

National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency,  
Mail Code 8623P, 1200 Pennsylvania Avenue, NW, Washington, D.C. 20460 USA

**Abstract.** Increased nutrient concentrations in streams and rivers have altered biological structure and function. Manipulative studies have provided insights into different mechanisms by which changes in nutrient concentrations influence aquatic biota, but these studies are limited in spatial scope and in their quantification of nutrient effects on aggregate measures of the invertebrate assemblage. Observational data provide a complementary source of information to manipulative studies, but these data must be analyzed such that the potential effects of spurious correlations are minimized. Propensity scores, a technique developed to analyze human health observational data, are applied here to estimate the effects of increased nutrients on the total taxon richness of stream invertebrates in a large observational data set collected from the western United States. The analysis indicates that increases in nutrient concentration are strongly associated with and cause decreases in invertebrate richness in large, but wadeable, open-canopied streams. These decreases in invertebrate richness were not mediated by periphyton biomass, a commonly proposed mechanism by which nutrients influence invertebrates. In smaller, closed-canopied streams, increases in nutrients were associated with small increases in total richness that were not statistically significant. Using propensity scores can greatly improve the accuracy of insights drawn from observational data by minimizing the potential that factors other than the factor of interest may confound the results.

**Key words:** confounding variables; macroinvertebrate; nitrogen; nutrients; phosphorus; propensity scores; streams; total taxon richness.

### INTRODUCTION

Human activities have greatly intensified the release of nitrogen and phosphorus in the environment (Vitousek et al. 1997, Smil 2000, Bennett et al. 2001), and these nutrients have altered species richness in many ecosystems (Tilman 1987, Valiela 1992, Dodson et al. 2000). In streams and rivers increased nutrients have markedly changed stream biological community structure and function. Increased nutrient concentrations alter the composition of algal assemblages (Rosemond et al. 1993, Slavik et al. 2004) and can often increase algal abundance (Hillebrand 2002). Increased nutrient concentrations also have been shown to alter microbial assemblages (Gulis and Suberkropp 2004), which in turn, influence decomposition rates (Gulis et al. 2004).

Invertebrates occupy an important position in the stream biological community (Covich et al. 1999), and attributes of the invertebrate assemblage (e.g., composition, biological function, richness of specific taxa) are widely used to assess stream condition (Moss et al. 1987, Fore et al. 1996). The responses of herbivore inverte-

brates to increased nutrients have been examined closely via field and laboratory studies because of explicit linkages between these taxa and increased nutrient concentrations and periphyton density (Feminella and Hawkins 1995). However, relatively little work has been done to examine the effects of nutrients on aggregate measures of invertebrate assemblages. In one example of work examining these aggregate measures, Cross et al. (2006) observed that secondary production in forested streams increased with nutrient enrichment. They also noted that much of this increase in production was associated with short-lived chironomid species (Cross et al. 2005). Aggregate measures, such as total taxon richness, are frequently used to interpret invertebrate data for biological assessment, and thus, current understanding of the effects of increased nutrients on invertebrate assemblages is incomplete in key areas needed to inform management actions.

Manipulative studies have contributed much to our understanding of nutrient effects in streams, but each individual study can only examine a limited range of conditions and endpoints. As a result, the current knowledge base provides uneven coverage of the linkages between invertebrates and nutrients with regard to different types of stream and with regard to different attributes of the assemblage structure. For example, many manipulative studies, particularly those based on

Manuscript received 19 September 2008; revised 13 February 2009; accepted 16 April 2009. Corresponding Editor: E. H. Stanley.

<sup>1</sup> E-mail: yuan.lester@epa.gov

whole stream manipulation (e.g., Elwood et al. 1981, Slavik et al. 2004, Sabater et al. 2005, Cross et al. 2006, Gafner and Robinson 2007) or artificial channels (e.g., Hart and Robinson 1990, Quinn et al. 1997), alter the nutrient concentrations and/or light availability in a single stream, and these results are difficult to generalize. Meta-analyses (e.g., Feminella and Hawkins 1995, Hillebrand 2002) have provided useful syntheses of the accumulated manipulative studies, but the general applicability of these analyses is still limited by the locations of the studies included in the analysis. As mentioned earlier, existing manipulative studies also provide limited insights into the effects of nutrients on the entire invertebrate assemblage. Those studies that consider effects on invertebrates often focus on a few key grazer taxa (e.g., snails, Rosemond et al. 1993; mayflies, Wellnitz et al. 1996; caddisflies, Hart and Robinson 1990) because these taxa are directly influenced by changes in the periphyton food base. Only a few studies (e.g., Quinn et al. 1997, Cross et al. 2006) have considered the effects of increased nutrients on aggregate measures of the invertebrate assemblage.

Observational data, collected from many streams spanning different typologies and different nutrient concentrations, can provide information that can complement knowledge gained from manipulative studies. Insights gained from analyzing these data can potentially help generalize findings from manipulative studies and highlight areas in which further experimentation would be fruitful. However, consistent associations between aggregate invertebrate assemblage attributes and nutrient concentrations have not been observed. For example, increased nutrient concentrations have been associated with both increases in total invertebrate taxon richness (e.g., Clenaghan et al. 1998, Heino et al. 2003) and decreases in richness (e.g., Miltner and Rankin 1998, Roy et al. 2003, Bergfur et al. 2007, Wang et al. 2007). Other studies found no significant associations between invertebrate richness and nutrient concentrations (e.g., Harding et al. 1999, Yuan and Norton 2003, Niyogi et al. 2007). The inconsistency of the conclusions drawn from observational studies may stem from limitations inherent in observational data. That is, demonstrating that observed associations in observational data represent causal relationships is often difficult because associations can be confounded by factors that covary with the nutrient concentrations.

Determining whether an observed phenomenon represents a cause-effect relationship requires that one compare conditions that can be observed (e.g., macroinvertebrate assemblage composition of a particular stream with 1 mg/L total nitrogen) with conditions that cannot be observed (e.g., macroinvertebrates in the same stream at the same moment in time with 0.1 mg/L nitrogen). Because this second case is unobservable, it is counter to fact, or counterfactual (Maldonado and Greenland 2002). Methods for quantifying causal effects provide different approaches for approximating the

unobservable, counterfactual case. Randomized, controlled experiments estimate the counterfactual case by replicating samples and then randomly assigning treatments (e.g., elevated nutrient concentrations) to a subset of the samples. The process of replication creates samples that are as close to identical as feasible and that differ only in whether a treatment is applied. Then, randomized assignment of treatments further protects against any remaining systematic differences between treated and control samples.

In observational data, treatments (e.g., different nutrient concentrations) have already been assigned by an unknown, nonrandom process, and so, to approximate the counterfactual case, we must identify samples that are similar to the sample of interest with regard to covariate distributions, but differ with regard to the factor of interest. For example, we must identify groups of streams that differ in their nutrient concentrations, but are similar with respect to other observed environmental factors. If only a single factor (e.g., agricultural land use) covaries with the factor of interest, we could simply stratify the data set by this factor, splitting the data set into groups with similar values. However, this approach rapidly becomes impractical as the number of factors increases. Propensity functions (Rosenbaum and Rubin 1983, Rosenbaum 2002, Imai and Van Dyk 2004) summarize the contributions of all known covariates as a single parameter. A propensity function is defined as the conditional probability of a multivariate treatment (e.g., different nutrient concentrations), given values of known covariates. This conditional probability can be characterized by a single parameter, referred to here as the propensity score, which is the mean expected value of the treatment. For example, observed nutrient concentrations can be modeled as a function of covariate values using regression analysis, and the predicted mean nutrient concentration in each stream is the propensity score. Then, stratifying by propensity score effectively splits the data set into groups with similar covariate distributions. Once the data set is stratified, causal effects of nutrients can be more confidently estimated within each group because distributions of other covariates are similar.

Propensity scores have been used frequently in epidemiological (e.g., Joffe and Rosenbaum 1999), sociological (e.g., Smith 1997), and economics studies (e.g., Dehejia and Wahba 2002), but thus far, have not been used to address ecological questions. Here, I use propensity scores to estimate the effects of increased nutrient concentrations on benthic invertebrates in small streams of the western United States. More specifically, I quantify the degree to which increased nutrients alter the total taxon richness of the benthic invertebrate assemblage. To examine a possible pathway by which nutrients can influence invertebrate richness, I also consider whether increased nutrient concentrations cause increased abundances of periphyton and invertebrate grazers.

## METHODS

*Data collection*

Regional scale data were collected by the U.S. Environmental Protection Agency Environmental and Monitoring Assessment Program at randomly selected Wadeable stream reaches across 12 states in the western United States in the summers of 2000 to 2002 (Stoddard et al. 2006). Extensive biological, physical habitat, chemical, and landscape-scale measurements were collected at each sampled site (Peck et al. 2008).

Identifying variables that covaried with nutrient concentration was an important preliminary step for this analysis. I identified these variables by examining bivariate scatter plots and correlation coefficients between total nitrogen and each of the different candidate variables. Variables that exhibited any evidence of a linear or nonlinear relationship (assessed qualitatively using scatter plots) were retained for the analysis.

Here, for conciseness, I only provide field collection methods for those variables used in the analysis. Stream water samples were collected at each site using a 4-L cubitainer and sent to a central laboratory for analysis (Peck et al. 2008). Nutrient concentrations were quantified as total N (TN) and total P (TP) by persulfate oxidation and colorimetry. Chloride ( $\text{Cl}^-$ ), sulfate ( $\text{SO}_4^{2-}$ ), ammonium ( $\text{NH}_4^+$ ), and bicarbonate ( $\text{HCO}_3^-$ ) ion concentrations were also measured in the collected sample using an ion chromatograph (U.S. EPA 2004). Sampling crews measured instantaneous stream temperature at the site at the time of sampling, and estimated percentage sand and fine substrate in the stream reach (SED [sediment]) by computing the proportion of particles that was  $<2$  mm in diameter, based on 105 systematic observations per site (Stoddard et al. 2006). Percentage open canopy cover was estimated from four measurements using a convex spherical densiometer collected at the middle of the stream channel (Peck et al. 2008). Sampling crews characterized riparian agricultural disturbance at each site by recording whether row crops or pasture existed in the riparian zone at the ends of 11 transects across the sampled reach. These observations were then summarized as a riparian disturbance index by counting the number of different types of disturbance and weighting by their proximity to the stream (Peck et al. 2008). Average annual precipitation (centimeters) at each site was estimated from long-term climatic summaries (Daly et al. 2001).

Several variables extracted from map data were also available in the data set. These variables included the geographical location of the site (decimal degrees latitude and longitude), watershed area ( $\text{km}^2$ ), and elevation (m). Land cover information from NLCD (National Land Cover Data; Vogelmann et al. 2001) based on satellite data at 30-m pixel resolution was aggregated into two major classes (agriculture and urban land use) known to influence nutrient loading.

Sites with urban land use in the catchment were sampled infrequently in this survey and were therefore difficult to model statistically. So, I omitted 13 sites with catchment urban land use  $>5\%$ . An estimate of grazing intensity was computed as a weighted product of land ownership, land cover, topography, slope, and proximity to water (see Appendix: Table A1; R. Comeleo, *personal communication*).

At each sampled reach, 11 equally spaced transects were established for collecting biological samples. Periphyton samples were gathered systematically from the left, center, and then right sides (looking downstream) of each successive transect after randomly assigning the collection location (left, center, or right) for the first transect. For sampling locations dominated by erosional habitat, periphyton was collected from  $12\text{ cm}^2$  of hard substrate by scrubbing and scraping for 30 s into a funnel, and then rinsing into a 500-mL bottle. For sampling locations dominated by depositional habitats, the top 1 cm of sediment from a  $12\text{-cm}^2$  plot was collected with a syringe. Samples from all transects were combined into a single composite sample; 25 mL of the composite sample was filtered through a glass fiber filter, and analyzed for chlorophyll *a* and ash-free dry mass. Macroinvertebrates were collected at the same time as periphyton with a D-frame kicknet ( $\sim 500\text{-}\mu\text{m}$  mesh) from  $0.09\text{-m}^2$  sampling plots on each transect ( $0.99\text{-m}^2$  total area). Invertebrate samples were composited, sieved, and preserved with 95% ethanol to a final concentration of  $\sim 70\%$ . In the laboratory, samples were spread on a gridded pan, and organisms were picked from randomly selected grid squares until at least 500 organisms were collected. Each organism was then identified to the lowest possible taxonomic level (usually genus).

A total of 827 distinct sites with complete biological and environmental data were available for analysis. Data used in this analysis are available online.<sup>2</sup>

*Statistical analysis*

I considered three biological responses: chlorophyll *a* per unit area of stream bed ( $\text{chl } a$ ,  $\text{mg/m}^2$ ), invertebrate grazer/scrapper relative abundance, and total invertebrate taxon richness. Observed taxa were assigned to functional feeding groups using the North American invertebrate traits database assembled by Vieira et al. (2006). Then, the relative abundance of grazer/scrapper taxa (hereafter referred to as scrapers) was calculated for each sample. Total taxon richness was calculated as the number of distinct invertebrate taxa observed in each sample.

Distributions of water chemistry measurements ( $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{HCO}_3^-$ , and TN), annual precipitation, and  $\text{chl } a$  were highly skewed and log-transformed prior to

<sup>2</sup> ([http://www.epa.gov/owow/streamsurvey/web\\_data.html](http://www.epa.gov/owow/streamsurvey/web_data.html))

analysis.  $\text{NH}_4^+$  and TP were  $\log(x + 1)$ -transformed to avoid log-transforming zero values.

$\log(\text{TN})$  and  $\log(\text{TP})$  were highly correlated ( $r = 0.68$ ), and together with low concentrations of  $\text{NH}_4^+$  were assumed to affect stream biota via similar models of action, so I modeled only the effects of TN and assumed that TN quantified a composite gradient of nutrient enrichment. The question of whether stream periphyton is N- or P-limited has been considered in many studies (e.g., Tank and Dodds 2003). However, meta-analyses (Francoeur 2001, Elser et al. 2007) suggest that both phosphorus and nitrogen addition stimulate periphyton growth. Recent work on ecological stoichiometry also suggests that the ratios of available nitrogen, phosphorus, and carbon can influence the growth rates of periphyton consumers (Sterner and Elser 2002, Evans-White and Lamberti 2006, Liess and Hillebrand 2006), but the relationships between these elemental ratios and aggregate invertebrate measures have not been considered. So, across the wide range of streams included in this study, I assumed that both TP and TN influenced stream biota, and TN concentrations represented the effects of both of these nutrients.

I estimated relationships between biological responses and TN using two analytical approaches. First, I directly applied conventional multiple linear regression models and nonparametric generalized additive models (GAMs; Wood and Augustin 2002) to the full data set to estimate relationships between  $\log(\text{TN})$  and each of the different biological responses (i.e., "direct models"). Fourteen variables that covaried with  $\log(\text{TN})$  (elevation, grazing intensity index, longitude,  $\log(\text{annual precipitation})$ ,  $\log(\text{catchment area})$ ,  $\log(\text{Cl}^-)$ ,  $\log(\text{HCO}_3^-)$ ,  $\log(\text{SO}_4^{2-})$ , SED, stream temperature, percentage of the catchment in agricultural land use, percentage of the catchment in urban land, percentage of open canopy, and riparian agricultural disturbance) were included in each of these models to control for their effects. Multiple linear regressions assume that the relationships between explanatory variables and the response can be modeled as straight lines. GAMs provide a more flexible approach for modeling these relationships because only the smoothness of the relationship is initially specified (via the degrees of freedom assigned to each curve fit). Then, a nonparametric curve is fit to the observed data. I allowed three degrees of freedom for each nonparametric curve fit, which permitted a wide range of possible responses (ranging from linear to unimodal functional forms), while not overfitting the data.

Second, I used propensity score analysis to estimate relationships between biological responses and  $\log(\text{TN})$ . Propensity score analysis required four steps as follows: (1) estimate the propensity score, (2) stratify the data set using the propensity score, (3) check whether comparable streams are included within each stratum, and (4) estimate the effects of TN within each stratum. I estimated the propensity score by using a GAM to model  $\log(\text{TN})$  concentrations as the sum of nonpara-

metric functions of the 14 covariates (Woo et al. 2008). Then, model-predicted mean values of  $\log(\text{TN})$  at each site were the propensity scores for each site (Imai and Van Dyk 2004).

I stratified the data set by dividing it into six groups based on equally spaced percentiles of the propensity score. Thus, within each stratum, the range of propensity scores was only approximately one-sixth that of the full data set. The appropriate number of strata to use depends somewhat on the data set. As the number of strata increases, the number of samples within each stratum (and the statistical inferential power) decreases. However, as the number of strata increases, propensity scores within each stratum span a narrower range of values, and covariate distributions are more similar. Rosenbaum (2002) suggests that five strata are appropriate for most data sets. I repeated the analysis using four and eight strata, and compared the results.

I quantified the effectiveness with which stratifying by propensity scores identified comparable streams by calculating correlation coefficients between  $\log(\text{TN})$  and other covariates within each stratum. If covariate distributions for streams are more similar within a stratum than across the entire data set, then we would expect the magnitudes of correlation coefficients between  $\log(\text{TN})$  and each covariate to be smaller in each stratum compared to the same correlation coefficients calculated using the full data set. I calculated correlation coefficients between  $\log(\text{TN})$  and the other covariates within each stratum and compared the maximum absolute values of these correlation coefficients across the six strata with correlation coefficients computed for the same pairs of variables using the entire data set.

I estimated the effects of TN on biological responses (chl *a*, scraper relative abundance, and total taxon richness) by fitting multiple linear regressions within each stratum. As suggested by Imai and Van Dyk (2004), all covariates used in the direct model were also included in models within each stratum. The propensity score for each site was independent of the choice of biological response so the same stratification was used for all three biological responses.

The average effect of  $\log(\text{TN})$  on each biological response for the entire data set (i.e., the average value of the regression coefficients) was computed as the average of coefficients of each stratum, weighted by the number of samples within each stratum. I also compared regression coefficient values across different strata for each biological response.

Structural equation models (SEM) can test whether observations are consistent with more complex models (Shipley 1999), so I used a simple SEM network to test the hypothesis that effects of  $\log(\text{TN})$  on total invertebrate richness were mediated by changes in chl *a* (i.e.,  $\log(\text{TN}) \rightarrow \text{chl } a \rightarrow \text{invertebrate richness}$ ). Within each stratum, I tested whether the observed covariance structure of these three variables was consistent with the hypothesized structure. Each model was assessed by

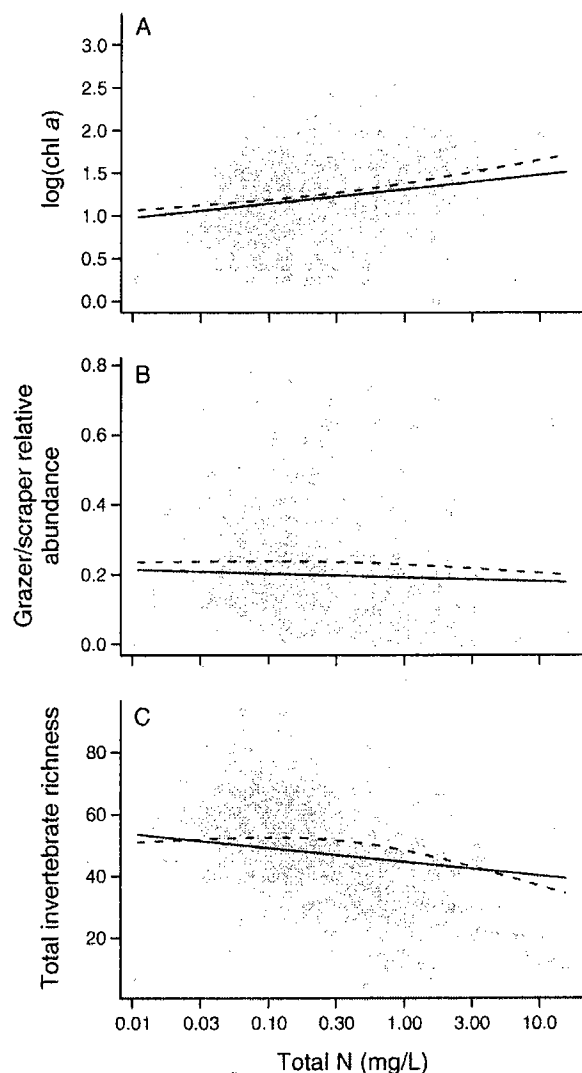


FIG. 1. (A) Log-transformed chlorophyll *a* (originally measured as  $\text{mg}/\text{m}^3$ ), (B) grazer/scrapper relative abundance, and (C) total invertebrate richness (number of taxa) plotted vs. total nitrogen (TN), shown on a log scale. Solid lines are relationships between TN and each biological response estimated by the direct multiple linear regression model. Dashed lines are the generalized additive model (GAM) estimate of the relationship.

using a chi-square statistic to test the null hypothesis that the observed covariance structure was consistent with a covariance structure generated from the hypothesized model and that any differences could be attributed to a multivariate normal sampling variability. A significant finding indicated that the null hypothesis was false, and the hypothesized model did not fully account for the observed covariances. Such a finding would suggest that TN affects invertebrate richness by a pathway that does not include chl *a*.

The statistical software R (R Development Core Team 2008) was used for all of the calculations, and the

structural equation modeling package available in R (sem) was used for the SEM.

## RESULTS

Across all sites, chl *a* was positively associated with  $\log(\text{TN})$ , scraper relative abundance showed no relationship, and total taxon richness was negatively associated with  $\log(\text{TN})$  (Fig. 1). In the direct multiple linear regression model, three variables ( $\log(\text{TN})$ , catchment agricultural land use, and  $\log(\text{Cl}^-)$ ) were significantly positively associated with chl *a* ( $P < 0.05$ ), while percentage substrate sand/fines and  $\log(\text{annual precipitation})$  were negatively associated ( $P < 0.001$ ). The model only accounted for 16% of the observed variance in chl *a*. Even after controlling for covariates in the model by fixing them at their mean values, the slope of the predicted mean relationship between  $\log(\text{TN})$  and chl *a* was qualitatively similar to the mean trend suggested by the data in the bivariate scatter plot (solid line in Fig. 1A). The relationship estimated from the nonparametric GAM exhibited only a slight deviation from the linear model (dashed line in Fig. 1A), and accounted for slightly more variance than the linear model ( $R^2 = 0.19$ ).

Scraper relative abundance was not associated with  $\log(\text{TN})$  (Fig. 1B). Instead, percentage substrate sand/fines and  $\log(\text{Cl}^-)$  were negatively associated with scraper abundance ( $P < 0.001$ ) in the direct multiple linear regression model, while grazing intensity index was positively associated ( $P < 0.05$ ). Slopes of the predicted mean relationships between  $\log(\text{TN})$  and chl *a* were very close to zero for both the multiple linear model and GAM (solid and dashed lines in Fig. 1B). The multiple linear regression model accounted for 13% of observed variance in scraper relative abundance, whereas the GAM accounted for 16%.

In the direct multiple linear regression model for total invertebrate richness, six variables ( $\log(\text{TN})$ , percentage substrate sand/fines, percentage open canopy, longitude,  $\log(\text{SO}_4^{2-})$ , and  $\log(\text{Cl}^-)$ ) were negatively associated with richness ( $P < 0.05$ ), and  $\log(\text{HCO}_3^-)$  and  $\log(\text{annual precipitation})$  were positively associated ( $P < 0.001$ ). This model accounted for 49% of observed variance in total richness. After using the regression model to control for the effects of covariates, the slope of the predicted mean relationship between  $\log(\text{TN})$  and invertebrate richness was much less than suggested by the data in the bivariate scatter plot (solid line in Fig. 1C), indicating that controlling for covariates reduced the strength of the estimated effect of  $\log(\text{TN})$  on total richness. The nonparametric relationship between  $\log(\text{TN})$  and total richness estimated by the GAM decreased more strongly at higher TN concentrations than at lower concentrations (dashed line on Fig. 1C). Qualitative estimates of the slopes for the low and high TN sections of the nonparametric curve were  $m = 1.4$  and  $-12.0$ , respectively, where  $m$  is the change in total richness per unit change in the  $\log(\text{TN})$  concentration.

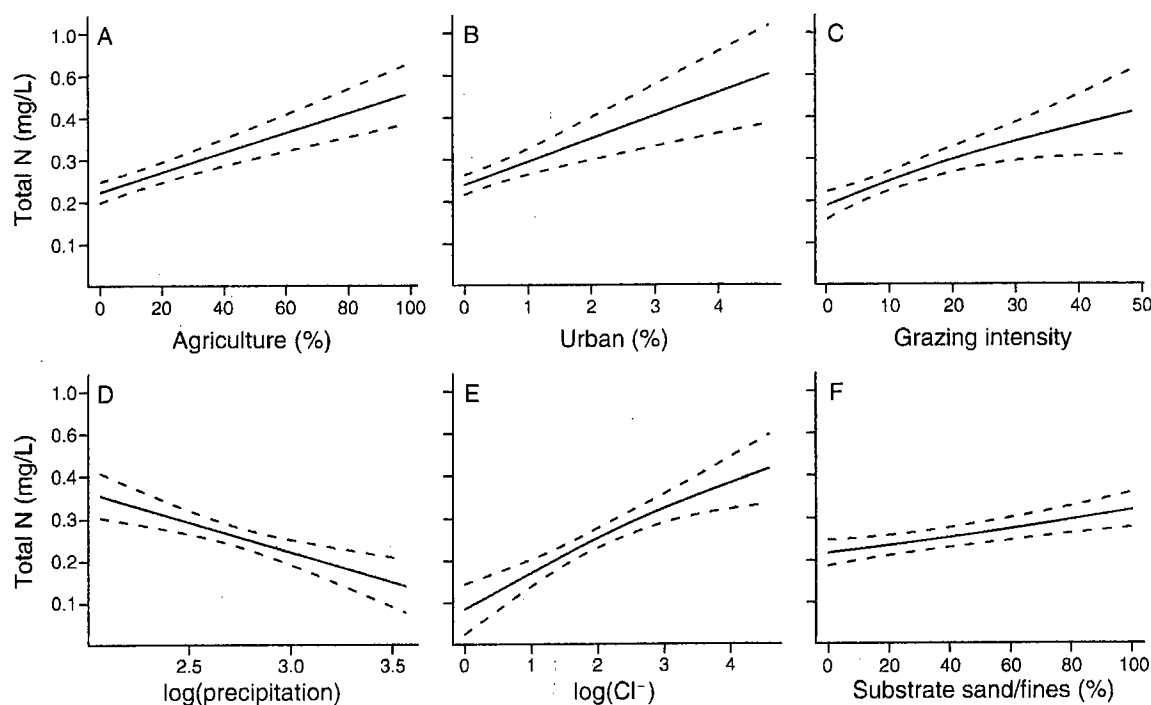


FIG. 2. Examples of relationships estimated between covariates and TN using GAM. Solid lines are mean relationships; dashed lines show 95% confidence intervals. Log-transformed precipitation was originally measured as mm/yr; log-transformed  $\text{Cl}^-$  was originally measured as  $\mu\text{eq/L}$ .

The GAM accounted for 52% of observed variance in total richness.

The GAM used to estimate the propensity score accounted for nearly 70% of observed variability in TN. Most of the nonparametric relationships between covariates and  $\log(\text{TN})$  were linear or nearly linear (Fig. 2).  $\log(\text{TN})$  increased with increases in catchment agricultural and urban land use and grazing intensity (Fig. 2A–C).  $\log(\text{TN})$  was also positively associated with  $\log(\text{Cl}^-)$  and substrate percentage sand/fines (Fig. 2E, F), and negatively associated with  $\log(\text{annual precipitation})$  (Fig. 2D).  $\log(\text{TN})$  increased with increased elevation and riparian agricultural disturbance as well (plots not shown). Note that the specific details of the propensity model fit were relatively unimportant compared to the assessment of how well the propensity scores estimated from the model controlled for covariance.

Strata defined by propensity scores varied in the range of propensity scores each encompassed (example for six strata shown in Fig. 3). Data were densest and the width of the strata narrowest in the middle of the distribution, whereas strata on the edges of the distribution (strata 1, 5, and 6) spanned larger ranges of propensity scores. Within each stratum, observed TN still varied by one or two orders of magnitude (Table 1). TP and  $\text{NH}_4^+$  increased across the strata as well (Table 1).

In the full data set,  $\log(\text{TN})$  was linearly correlated with most of covariates considered with  $|r| > 0.45$

(Table 2). Two candidate variables (elevation and percentage catchment urban land use) were not strongly correlated with  $\log(\text{TN})$ , but were included in the propensity model because of their known associations with invertebrate richness and TN (see *Discussion*). In the strata defined by the propensity scores, correlations between  $\log(\text{TN})$  and all modeled covariates were much weaker (Table 2). For example, across six strata ( $n_s = 6$ ), the maximum magnitude of the correlation coefficient between  $\log(\text{TN})$  and grazing intensity index was only

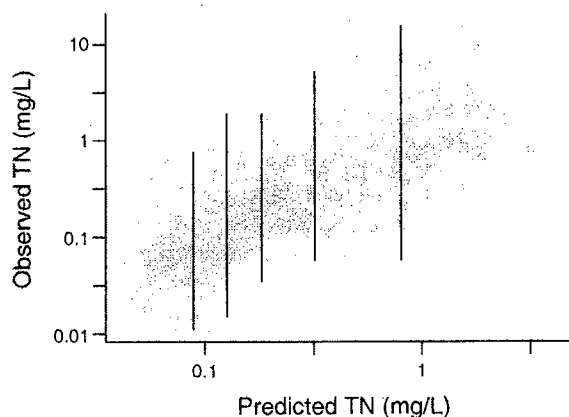


FIG. 3. Observed values of TN vs. GAM regression predictions of TN. These regression predictions are the propensity scores. Solid vertical lines show the strata boundaries for six strata. Note the log–log scale.

TABLE 1. Mean, minimum, and maximum values for total nitrogen (TN), ammonium ion ( $\text{NH}_4^+$ ), and total phosphorus (TP) in stream strata in the western United States defined by propensity scores.

Stratum	TN (mg/L)			$\text{NH}_4^+$ (mg N/L)			TP (mg/L)		
	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
1	0.080	0.011	0.576	0.006	0.001	0.039	0.008	0.000	0.065
2	0.120	0.015	0.744	0.007	0.000	0.022	0.017	0.000	0.082
3	0.188	0.035	1.898	0.010	0.001	0.134	0.025	0.000	0.153
4	0.284	0.060	1.798	0.011	0.001	0.211	0.041	0.001	0.203
5	0.636	0.058	5.113	0.024	0.001	0.389	0.126	0.002	5.418
6	1.444	0.134	15.625	0.063	0.004	1.148	0.261	0.003	7.633

Notes: Propensity scores characterize the mean probability of a stream reach being exposed to a certain concentration of total N, conditioned by other measured covariates. See *Methods: Statistical analysis* for details regarding estimation of propensity scores. Strata are ordered with respect to increasing propensity scores.

0.19 (compared to 0.64 for the full data set). The strengths of correlation between other covariates and  $\log(\text{TN})$  were reduced by similar amounts. The strength of correlation between elevation and  $\log(\text{TN})$  was somewhat greater after stratification than in the full data set, but in all stratification schemes considered, this moderately strong correlation occurred only in the last stratum. Across the remaining strata, the maximum correlation between elevation and  $\log(\text{TN})$  ranged only from 0.08 to 0.26 (Table 2). Overall, the number of strata used did not substantially change the effectiveness of the stratification: in all cases that were considered, the degree to which  $\log(\text{TN})$  was correlated with covariates was greatly reduced after stratification. Thus, regression estimates of the effects of TN on biological responses within each stratum could be more confidently characterized as causal effects.

Propensity scores and the strata based on these scores roughly defined a gradient of streams ranging from small, shaded streams with small amounts of fine substrate (stratum 1) to larger, open canopy streams with large amounts of fine substrate (stratum 6) (Fig. 4; also see Plate 1). Human disturbance in the catchments

for streams in strata 5 and 6 also was more intense, with greater catchment percentages of row-crop agriculture and greater grazing intensities.

Distributions of the biological responses varied across strata. These variations were particularly noticeable for total invertebrate richness, which decreased steadily from stratum 1 to stratum 6 (Fig. 5). Mean values of chl *a* increased slightly from stratum 1 to 6, while scraper relative abundance did not exhibit any strong patterns.

Using propensity score analysis, estimates of the average magnitude of the effect of  $\log(\text{TN})$  on chl *a* (i.e., the average value of linear regression coefficients across all strata) were similar regardless of the number of strata considered, ranging from an increase of 0.18 to 0.21 of  $\log(\text{chl } a)$  ( $\text{mg}/\text{m}^2$  of stream bed) per unit change in  $\log(\text{TN})$  (Table 3). These effect estimates were similar to those estimated by the direct multiple linear regression model, and all effect estimates were statistically significant ( $P < 0.05$ ). Increased TN was associated with small decreases in scraper relative abundance, but all of these decreases were not statistically significant. The average magnitude of the effect of  $\log(\text{TN})$  on total invertebrate richness depended somewhat on the num-

TABLE 2. Comparison of correlation coefficient between TN and different covariates.

Covariate	$r_f$	Max $ r_s $		
		$n_s = 4$	$n_s = 6$	$n_s = 8$
Elevation	-0.21	0.40 (0.08)	0.31 (0.21)	0.26 (0.26)
Grazing intensity index	0.64	0.23	0.19	0.23
$\log(\text{annual precipitation})$	-0.51	0.29	0.25	0.29
$\log(\text{catchment area})$	0.51	0.10	0.23	0.10
$\log(\text{Cl}^-)$	0.64	0.14	0.12	0.28
$\log(\text{HCO}_3^-)$	0.56	0.14	0.21	0.22
$\log(\text{SO}_4^{2-})$	0.64	0.14	0.07	0.20
Longitude	0.63	0.19	0.24	0.14
Percentage of catchment in agriculture	0.61	0.38	0.29	0.24
Percent of catchment in urban	0.34	0.16	0.12	0.30
Percentage open canopy	0.49	0.15	0.11	0.11
Percentage of sand/fines in substrate	0.65	0.15	0.26	0.22
Riparian agricultural disturbance	0.53	0.10	0.21	0.14
Stream temperature	0.48	0.11	0.20	0.21

Notes: For the column headings,  $r_f$  is the correlation coefficient computed for the full data set, and  $r_s$  is the correlation coefficient computed within each stratum. Max  $|r_s|$  is the maximum absolute value of  $r_s$  across all strata, and  $n_s$  is the number of strata. Numbers in parentheses in the first row are maximum absolute correlation coefficients computed after omitting the stratum with the highest mean TN concentration.

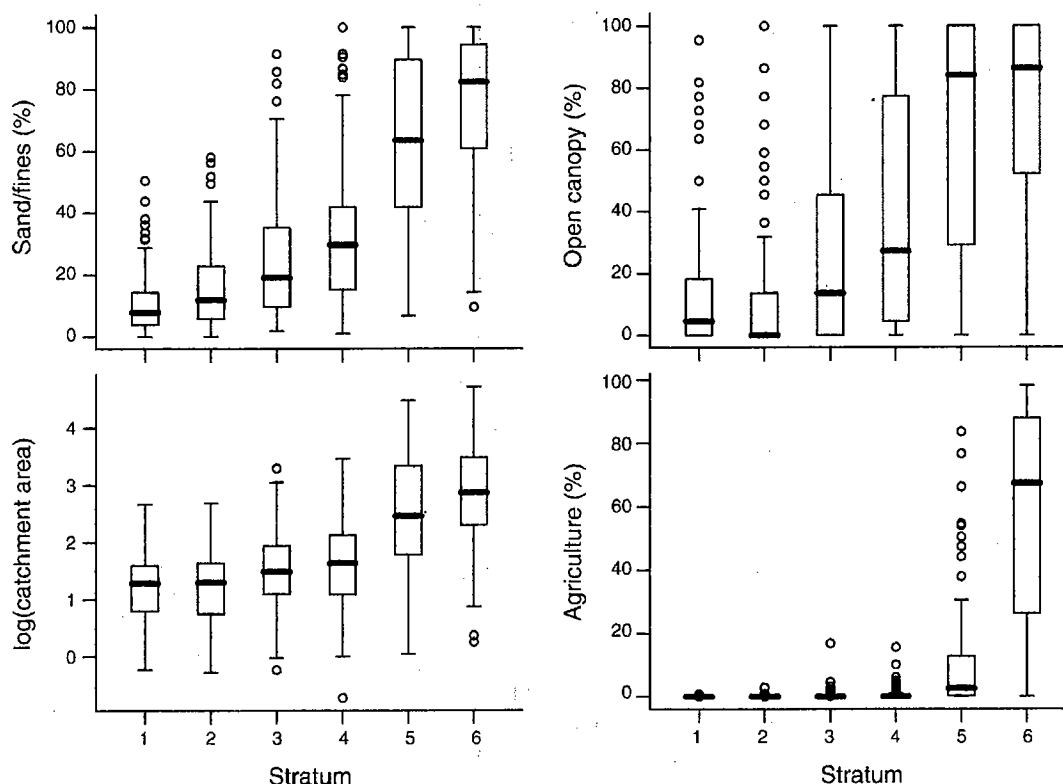


FIG. 4. Examples of variations in covariate (percentage sand/fines in substrate, percentage of open canopy, log-transformed catchment area [originally measured in  $\text{km}^2$ ], and percentage of the catchment in agriculture) distributions across strata defined by propensity scores. Boxes indicate the 1st quartile, median, and 3rd quartile values, respectively. Vertical lines extend to 1.5 times the interquartile range, and open circles indicate values that are outside this range.

ber of strata considered. With four strata, a decrease of 3.6 taxa per unit increase in  $\log(\text{TN})$  was estimated, but for six and eight strata, the estimated effect sizes were decreases of 1.9 and 2.1 taxa per unit increase in  $\log(\text{TN})$ , respectively. The direct multiple linear regression model estimated the greatest decrease of taxa richness with increased  $\log(\text{TN})$  ( $-4.5$  taxa per unit increase in  $\log(\text{TN})$ ).

The effects of  $\log(\text{TN})$  on total invertebrate richness varied substantially across strata defined by propensity scores (example for  $n_s = 6$  shown in Fig. 6). In the first three strata, increases in  $\log(\text{TN})$  were associated with small increases in total invertebrate richness. In the next three strata (strata 4–6), statistically significant decreases of  $-6.0$  to  $-9.5$  taxa per unit increase in  $\log(\text{TN})$  were observed.

Estimated effects (i.e., linear regression coefficients) of TN on all three biological responses across different strata (for  $n_s = 6$ ) are summarized in Fig. 7, where estimated effect sizes are plotted against the mean value of predicted  $\log(\text{TN})$  within each stratum. As observed previously, total invertebrate richness increased slightly with increased TN in the first three strata, but total invertebrate richness decreased with increased TN for all subsequent strata (Fig. 7C). TN had virtually no effect on chl  $a$  in the first stratum, but in the next two strata,

chl  $a$  increased slightly (but not significantly) with increased TN (Fig. 7A). Significant positive associations between chl  $a$  and  $\log(\text{TN})$  were observed in stratum 4 (regression coefficient  $[b] = 0.60$ ; standard error  $[\text{SE}] = 0.16$ ) and stratum 6 ( $b = 0.30$ ,  $\text{SE} = 0.10$ ). The effects of TN on scraper relative abundance varied slightly across strata, but none of the regression coefficients were statistically significant (Fig. 7B).

Observed covariance structure within each stratum differed significantly from the covariance structure predicted by the simple structural equation model in all strata except for strata 1–3 (Table 4). When applied to the full data set, the null hypothesis was also rejected strongly ( $P < 0.001$ ). Thus, in strata 4–6 and across the entire data set, observations of  $\log(\text{TN})$ , chl  $a$ , and invertebrate richness were not consistent with the model that decreases in invertebrate richness were mediated by changes in chl  $a$ .

#### DISCUSSION

Increased nutrients have been associated with alterations in species richness in a variety of ecosystems and for a variety of taxonomic groups. Long-term nitrogen loading decreases plant species richness (Tilman 1987) and insect richness (Haddad et al. 2000) in grasslands. In lakes, a unimodal relationship is often observed between

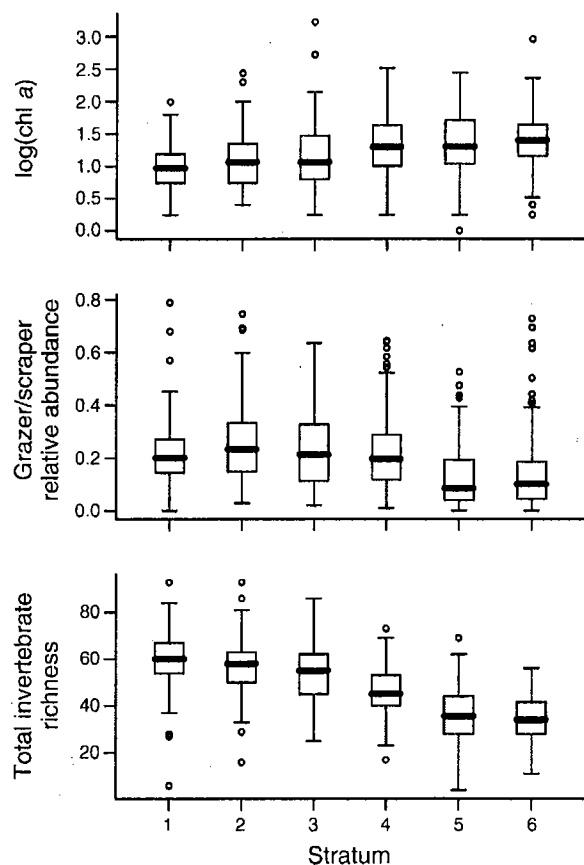


FIG. 5. Distributions of biological responses across propensity score strata. Box plots are as described in Fig. 4. Log-transformed chlorophyll *a* was originally measured as  $\text{mg}/\text{m}^2$ .

primary productivity and species richness (Dodson et al. 2000), and differences in nutrient availability may cause many of these variations in primary productivity (Smith 1979). In estuaries, increased nutrient loading has also been associated with decreased benthic invertebrate richness (Valiela et al. 1992). My analysis complements and extends existing knowledge on effects of nutrients on species richness, indicating that increased nutrients in certain types of streams cause significant decreases in invertebrate richness. However, before we can consider the present results in the context of existing knowledge, we must consider whether we can interpret the estimated effects as cause-effect relationships.

#### Observational data and cause-effect relationships

Estimating cause-effect relationships from observational data is difficult because it is often impossible to determine whether the effect that is observed is caused by the factor of interest (e.g., increased nutrients) or whether it is caused by other factors that are correlated with the factor of interest. Manipulative studies provide stronger evidence of cause-effect relationships than observational studies because treatments (and the magnitude of treatments) are assigned randomly; and therefore, potential confounding variables are randomly assigned as well (Rosenbaum 2002), and on average, distributions of confounding variables are balanced across treatment and control samples. Propensity scores provide an analytical means of directly controlling for many different potential confounding variables in observational data. Here, I used propensity scores to control for a variety of confounders, including in-stream conditions, land cover, and natural template (Table 2). By reducing the strength of the correlations between covariates and  $\log(\text{TN})$ , the potential for these covariates to confound estimated relationships was reduced. Furthermore, the range of variable values that had to be represented by linear regression was reduced, which reduced the potential effects of model misspecification.

The propensity score approach can only control for variables for which data are available, and so unobserved variables could still potentially confound the present results. I tried to minimize the effects of unobserved variables by including many different types of covariates in the propensity model. Including variables correlated with  $\log(\text{TN})$  that are known to directly influence invertebrate assemblage composition (percentage substrate sand/fines,  $\text{Cl}^-$ ,  $\text{HCO}_3^-$ , and  $\text{SO}_4^{2-}$ ) provides the most powerful means of controlling for confounders. Unfortunately, only limited data on these factors were available. Indeed, many important proximal environmental factors may not even be known, so I also included possible sources of in-stream factors (e.g., catchment land cover) that could influence invertebrate assemblages and were correlated with  $\log(\text{TN})$ . These variables have been shown to be strongly associated with invertebrate composition (e.g., Richards et al. 1996), and one possible pathway by which they influence invertebrate assemblages is via changes in nutrient concentration. However, catchment

TABLE 3. Average effect (i.e., linear regression coefficient, *b*) of TN on each biological endpoint across the entire data set, and the standard error of the estimate (SE).

<i>n<sub>s</sub></i>	Chlorophyll <i>a</i>		Grazer/scrapper relative abundance		Total taxon richness	
	<i>b</i>	SE	<i>b</i>	SE	<i>b</i>	SE
4	0.18	0.052	-0.018	0.016	-3.6	1.3
6	0.21	0.054	-0.017	0.017	-1.9	1.4
8	0.21	0.055	-0.025	0.017	-2.1	1.4
Direct model	0.17	0.049	-0.011	0.015	-4.5	1.3

Note: Results were tabulated for different numbers of strata (*n<sub>s</sub>*) used in propensity score analysis and for the direct model.

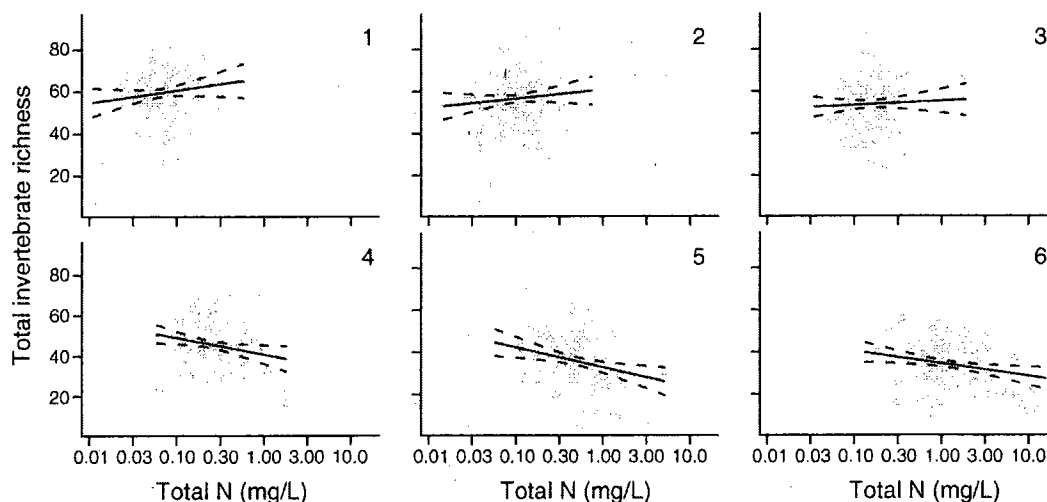


FIG. 6. Total invertebrate richness (number of taxa) vs. TN (note log scales) in different propensity score strata; panels are labeled by stratum number. Solid lines are mean estimated linear relationships; dashed lines show 95% confidence intervals.

land uses such as row crop agriculture are also associated with a number of other environmental factors (e.g., increased pesticides) that can influence invertebrate assemblage structure. By including these source variables in the propensity score model, I at least partially controlled for some of these other factors, strengthening the causal case for an effect of TN on the biological responses. Elevation and percentage catchment urban land use were only weakly associated with  $\log(\text{TN})$  but were included in the propensity model because others have documented their relationships with TN and with one or more of the biological responses (Roy et al. 2003, Vinson and Hawkins 2003). Ultimately, the propensity score model was limited by available data, so some potentially important factors, such as flow variability, were not included. However, the included covariates provided reasonably comprehensive coverage of the variables that I would expect to covary with TN and influence the biological responses considered here.

After stratifying by the propensity scores, correlation strengths between most confounding variables and  $\log(\text{TN})$  were greatly reduced, but elevation was still moderately correlated with  $\log(\text{TN})$  in certain strata. Scatter plots (not shown) indicated that the relationship between  $\log(\text{TN})$  and elevation was strongly heteroscedastic, with much higher variances of  $\log(\text{TN})$  observed at lower elevations compared to higher elevations. Consequently, the propensity score model was not able to effectively control for the confounding effects of elevation in all strata. However, it is likely that elevation exerted a weak influence on the estimated effects of TN for the following reasons. First, correlation strength between  $\log(\text{TN})$  and elevation was the strongest for  $n_s = 4$  and decreased to levels comparable with the other covariates when 6 and 8 strata were used (Table 2). Second, correlations between elevation and  $\log(\text{TN})$  only exceeded 0.3 in the last stratum for the  $n_s =$

4 and 6 cases, so the potential confounding effects of elevation were limited.

Estimates of the effects of TN on invertebrate richness were somewhat sensitive to the number of strata used (Table 3). Results calculated using four strata differed particularly from results calculated using six and eight strata. Correlation strength between  $\log(\text{TN})$  and percentage catchment agriculture remained at 0.38 for the four-strata case, compared to 0.29 and 0.24 for the other two cases (Table 2), and this difference may explain the different estimates of the biological effects of TN. Overall, it seems that the general guidance for using at least five strata is appropriate for this data set (Rosenbaum 2002).

One question one must consider when assessing post-stratification correlation strengths is whether they are low enough to reasonably assure that covariates do not bias the estimated effect of interest. Mathematically, if we modeled biological responses solely as a linear function of  $\log(\text{TN})$ , a correlation strength of zero between  $\log(\text{TN})$  and any possible covariate would guarantee that the covariate in question did not influence the estimation of  $\log(\text{TN})$  regression coefficient. However, attaining correlation strengths that approach this ideal is generally impractical. Conversely, even a covariate that is weakly correlated with  $\log(\text{TN})$  still retains some potential for biasing the regression coefficient estimate, especially if the covariate itself strongly influences the biological response. In general, though, covariates that are weakly correlated with the factor of interest have much less potential for biasing the estimated effect. Furthermore, in the present analysis, all covariates were included in the regression models fit for each stratum. Including covariates in the regression models provides a second level of insurance that analysis results are unbiased because covariates that strongly

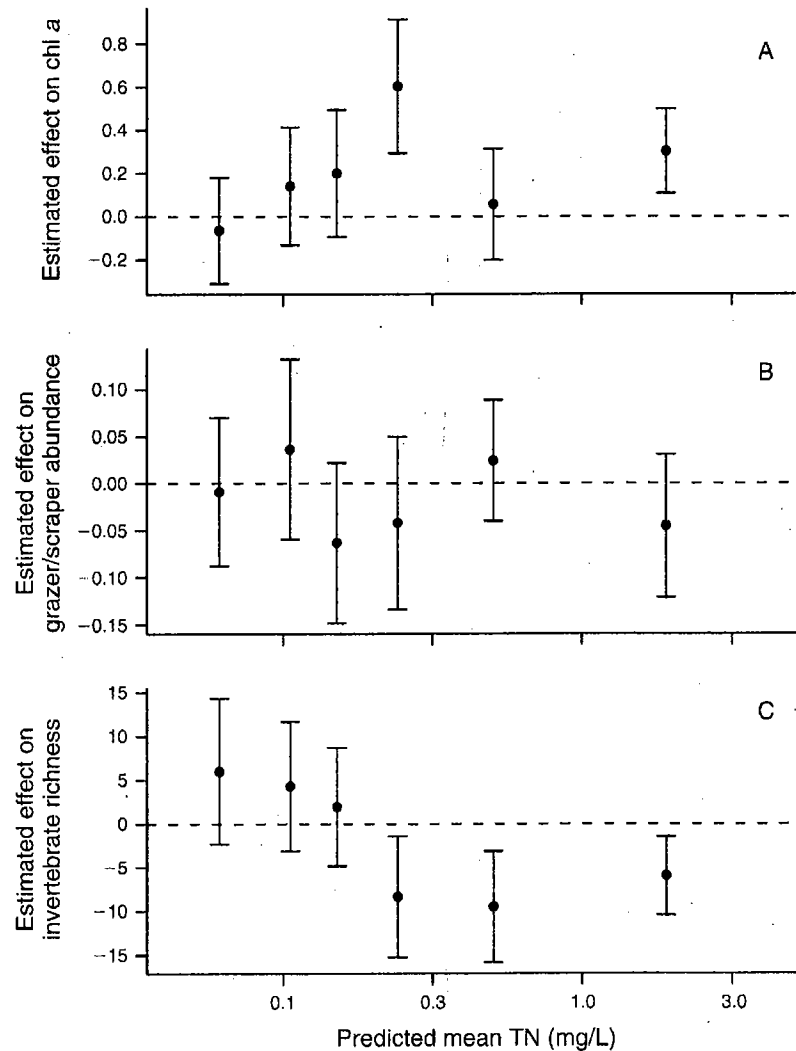


FIG. 7. (A) Estimated effects (i.e., linear regression coefficients) of TN on chl *a*, (B) grazer/scrapper relative abundance, and (C) total invertebrate richness. Error bars indicate 95% confidence intervals. Note the log scale for TN.

influence the biological response still can contribute to the model for observed biological responses.

Including confounding variables in a multiple linear regression (i.e., the direct model) is a common approach for controlling for the effects of confounding variables, but many of the assumptions inherent in this approach must be considered carefully when interpreting results. When using multiple linear regression, one assumes that the relationships between explanatory variables and responses are linear functions and that no interactions occur between explanatory variables. These assumptions can be slightly relaxed by adding additional nonlinear or interacting terms to the model equation, but doing so requires prior knowledge of which variables are likely to interact and/or which variables are nonlinearly related to the response. Generalized additive models (Wood and Augustin 2002) provide further flexibility for specifying the shape of response functions, and in the present

analysis, the GAM estimate of the relationship between  $\log(\text{TN})$  and invertebrate richness more closely approximated the relationships estimated with propensity score analysis (compare Fig. 1C and Fig. 6). Ultimately, though, all regression models are mis-specified to some

TABLE 4. Maximum-likelihood chi-square statistics for structural equation models fit in each stratum and for the direct model.

Stratum	$\chi^2$	<i>P</i>
1	1.0	0.32
2	0.3	0.58
3	1.8	0.10
4	8.0	0.005
5	17.5	<0.001
6	11.2	<0.001
Direct model	282.5	<0.001

Note: There is one degree of freedom for each stratum.



PLATE 1. Examples of streams from stratum 3 (left photo, Fall River, Colorado, USA) and stratum 4 (right photo, Cataract Creek, Colorado, USA). Photographs courtesy of Colorado Division of Wildlife.

degree and can provide misleading results, as observed here with the direct models for total invertebrate richness (Table 3). By partitioning the data set into strata, propensity score analysis greatly reduces the potential impact of model mis-specification (Imai and Van Dyk 2004). Confounding variables and the causal variable of interest are restricted in their range of values within each stratum, and covariates are restricted in the degree to which they are correlated with the causal variable of interest. Therefore, linear functions are more likely to provide appropriate representations of the relationships between variables and responses. Similarly, the potential strength of interacting effects is greatly curtailed by the restricted range of values.

Because the covariates included in this analysis are fairly comprehensive and because the models used are robust to model misspecification, the estimated relationships between biological responses and TN presented here can be regarded as reasonably reliable estimates of cause-effect relationships.

#### *Effects of increased nutrients on stream biota*

My analysis showed that increased TN concentrations in certain streams caused statistically significant decreases in invertebrate richness. These findings are consistent with previous observational studies that have reported strong associations between nutrients and richness (e.g., Wang et al. 2007), but as noted earlier, relationships between nutrient concentrations and invertebrate richness reported from observational studies have not been consistent. My analysis improves on past analyses of observational data by asserting that observed relation-

ships represent cause-effect relationships. Such strong causal evidence typically requires manipulative studies but in streams to date, few manipulative studies have examined the relationship between nutrient richness and invertebrate richness. However, in other ecosystems, the effects of nutrient manipulations on species richness have been reported. For example, in grasslands, long-term nutrient enrichment can reduce plant and insect richness (Tilman 1987, Haddad et al. 2000). Similarly, manipulations of nutrient concentrations in lakes have revealed negative associations between nutrient concentrations and phytoplankton richness (Dodson et al. 2000).

In streams, manipulative studies have focused primarily on the interactions between nutrient enrichment, periphyton, and invertebrate grazers and scrapers, and the agreement between my results and results reported from these studies was somewhat mixed. Causal relationships between increased TN and increased chl *a* identified in my analysis (Table 3) were broadly consistent with findings from manipulative studies (see Feminella and Hawkins 1995, Hillebrand 2002 for summaries). Many studies (Rosemond et al. 1993, Feminella and Hawkins 1995) have also shown that invertebrate grazer or herbivore abundance and biomass increase in response to nutrient addition, but I did not observe significant relationships between scraper relative abundance and chl *a* here. This finding was perhaps not too surprising, given the errors and simplifications associated with the assignment of functional feeding groups to different invertebrate taxa. Food choice for many invertebrates can depend on the availability of different food sources (e.g., Mihuc and Minshall 1995, McCutchan and Lewis 2002)

and can vary substantially over the life span of the organism (Merritt and Cummins 1996). Thus, the designation of feeding groups for the many different taxa collected in this study likely introduced errors that may have obscured relationships between scraper relative abundance and TN. In contrast, most manipulative studies focus specifically on one or two taxa whose feeding behavior is well established. The use of relative abundance to quantify the role of scraper taxa in the invertebrate assemblage may also have contributed to the lack of a significant relationship because abundance measurements do not represent potential differences in the grazing impact of large-bodied vs. small-bodied herbivores (Holomuzki and Biggs 2006). Biomass would have been a more appropriate measure for quantifying a scraper response to increased periphyton; however, collection of biomass data is costly, and it is rarely available in large-scale observational data sets. Collection of biomass data in large-scale synoptic surveys would greatly enhance our ability to verify and extend findings from manipulative studies.

#### *Variations in biological response across strata*

The response of both invertebrate richness and chl *a* to increases in TN varied substantially across different strata, and these variations may provide insights into the mechanisms by which changes in TN concentrations affect invertebrate richness. Stream types (*sensu* Frissell et al. 1986) varied within each stratum. Streams in strata 1–3 had small levels of substrate sediment and closed canopies, and drained relatively small catchment areas (Fig. 4), characteristics often attributed to headwater streams (Vannote et al. 1980). Conversely, streams in strata 4–6 had larger amounts of substrate sediment and open canopies, and drained larger catchments. Average nutrient concentrations also increased from stratum 1 to 6 (Table 1). These differences across strata likely influenced the responses of different biological factors to changes in TN. Concentrations of TN observed in this data set were not high enough to be directly toxic to invertebrates, and so, the effect of TN on invertebrates must be mediated by another factor that can be directly influenced by TN. Here, I consider periphyton growth and microbial activity as two possible mediating factors, and discuss possible mechanisms by which these factors could influence invertebrate richness differently in different strata.

Increases in nutrients increase periphyton biomass and alter the periphyton assemblage composition (Chételat et al. 1999), and both of these phenomena can then influence invertebrate richness. Increased periphyton biomass can provide additional food for certain invertebrates, potentially supporting a greater diversity of taxa. Conversely, decomposition of large amounts of periphyton biomass can episodically reduce dissolved oxygen, which could reduce invertebrate richness (Correll 1998). Chl *a* provided a direct, albeit imperfect, measure of periphyton biomass in my analysis, and SEM tests indicated that chl *a* did not

mediate the linkage between TN and invertebrate richness in strata 4–6. Thus, we can likely conclude that the mechanism in which increased periphyton biomass reduces invertebrate richness is not operative in the streams sampled in this data set. In contrast, SEM tests in strata 1–3 did not reject the possibility that increased chl *a* mediated the positive association between TN and invertebrate richness. However, all biological responses to TN in these strata were not statistically significant.

Increased TN can also alter periphyton composition, increasing the dominance of particular species (Chételat et al. 1999), and the concurrent decrease in periphyton species richness may cause decreases in invertebrate richness. Experiments in grasslands have shown that increased nutrients reduce plant and insect richness (Tilman 1987, Haddad et al. 2000). In these studies, changes in insect richness were best explained by changes in plant richness (rather than nutrient levels), suggesting that changes in insect richness were mediated by changes in plant richness. This same mechanism may explain negative associations between invertebrate richness and TN in strata 4–6.

Microbial (fungi and bacteria) activity is also severely limited by nutrient availability (Morris and Lewis 1992), and increases in nutrients can markedly enhance the growth of these organisms. Increased microbial activity can then alter invertebrate richness by enhancing the nutritional quality of detrital matter, by accelerating decomposition rates, and by causing direct mortality of certain invertebrates. Increased microbial activity may accelerate the conditioning of detrital matter for invertebrate consumption (Allan 1995), and this improved food quality may support a greater number of invertebrate taxa, as observed in the first strata 1–2 (Fig. 6). Manipulative studies support this idea, as Cross et al. (2006) suggested that increased abundances of short-lived taxa and secondary production in nutrient-enriched streams were due to improved detritus quality. Furthermore, the highly shaded, headwater streams included in strata 1–2 were very comparable to the streams in which Cross et al. (2006) performed their experiments. Further enhancement of microbial activity may further accelerate decomposition (Gulis et al. 2004) to the point at which potential food is decomposed more rapidly than it can be used by invertebrate consumers. As available food decreases, we would expect a decrease in invertebrate richness by competitive exclusion, a mechanism that has been proposed to explain negative associations between nutrient concentrations and invertebrate richness in lakes (Dodson et al. 2000). As noted earlier, rapid decomposition of periphyton biomass can also potentially episodically reduce dissolved oxygen, a phenomenon can cause mortality or increased emigration of invertebrates. We might expect these rapid decomposition rates in strata 4–6, where nutrient concentrations are higher and less riparian canopy cover

can potentially cause higher daytime water temperatures (Kaplan and Bott 1989).

Increased microbial activity can also influence invertebrate assemblages directly by increasing the potential for bacterial colonization of particular invertebrate taxa. Lemly (1998) and Fuller et al. (2004) observed that in streams with high bacteria concentrations, bacteria could coat the exposed gills of certain invertebrates. Subsequent laboratory tests then demonstrated that bacterial coatings markedly increased the mortality of certain invertebrate taxa (Lemly 1998). As noted earlier, very high nutrient concentrations and higher temperatures in streams of strata 4–6 provide conditions that are suitable for this phenomenon.

### Conclusions

Stream invertebrate richness has been associated with a wide variety of different anthropogenic and natural factors (e.g., Fore et al. 1996, Roy et al. 2003). The present analysis lends further evidence of the broad responsiveness of invertebrate richness to different environmental factors, as the direct regression model found that seven variables in addition to log(TN) were significantly associated with invertebrate richness. This broad responsiveness of invertebrate richness is one reason why richness is so widely used to assess stream biological condition (Moss et al. 1987, Kerans and Karr 1994); however, it complicates efforts to isolate the effects of nutrients because many of the factors that cause declines in invertebrate richness are also correlated with TN. The propensity score analysis introduced here controls for the confounding factors for which data were available and strengthens the case for inferring that observed associations reflect cause–effect relationships.

This analysis establishes a causal relationship between TN concentrations and invertebrate total richness, but elucidating the mechanism by which this effect occurs requires further study. In particular, manipulative studies in the more degraded streams that composed strata 4–6 for this study and increased attention to total invertebrate richness would be particularly useful. The existing emphasis on grazers in studies linking invertebrates to nutrient enrichment is understandable: the combination of bottom-up influences on periphyton abundance (via nutrient enrichment) and top-down controls (via grazing) provides a model system in which to verify and develop ecological theory (Hillebrand 2002). However, from the perspective of specifying appropriate management actions for excess nutrients in streams and rivers, these studies are of limited utility. Invertebrate richness is a key component of metrics commonly used to assess the biological condition of streams (Moss et al. 1987, Kerans and Karr 1994), and thus, quantifying the relationship between nutrients and other stressors and invertebrate richness would provide valuable information for managers seeking to identify target nutrient concentrations or criteria (Dodds and Welch 2000).

One key difference between the data I used in this analysis and most existing manipulative studies is the duration of the exposure to elevated nutrient concentrations. Here, streams in which observed TN concentrations were high likely experience chronically high nutrient concentrations, and these concentrations change only on the timescale of land use changes (i.e., years). Furthermore, the present analysis considers the relationship between TN and biological responses across different streams. In contrast, most manipulative studies elevate nutrient concentrations for only a short amount of time in the same stream. Cross et al. (2006) enriched their study stream for two years, but most other studies (e.g., Gafner and Robinson 2007) enrich for much shorter durations. The differences in study duration and the uncertainties inherent in a space-for-time switch greatly increase the uncertainty of identifying the correct mechanism by which the elevated nutrients decrease invertebrate richness. For example, Cross et al. (2006) observed that increased nutrients increased decomposition rates, but they speculated that a longer exposure to elevated nutrients would ultimately reduce the quantity of available organic matter in the stream. As noted earlier, streams with chronically high nutrient concentrations may have less available organic matter, which may contribute to reduced invertebrate richness. Slavik et al. (2004) observed drastic changes in stream biotic composition during their long-term nutrient enrichment, suggesting that the long-term mechanisms by which nutrients influence stream biota can differ substantially from those that are observed in short-term experiments.

Observational data provide an important source of information for better understanding relationships between human activities and ecological responses. These data can span a wide range of streams and include the full natural complexity of the studied ecosystem. As such, appropriately designed analyses of these data can complement and extend knowledge gained from manipulative studies, and help generate hypotheses for additional experiments. Propensity score analysis provides one particularly robust method for analyzing observational data. Combining the specific mechanistic insights gained from manipulative studies with the general trends that can be extracted from observational data ultimately can lead to a more complete understanding of the relationships between environmental factors and biological responses.

### ACKNOWLEDGMENTS

The author gratefully acknowledges comments from A. I. Pollard, L. Alexander, S. B. Norton, and two anonymous reviewers that greatly improved the manuscript. Citations included in the interactive conceptual model (<http://cfpub.epa.gov/caddis/icm/ICM.htm>) were particularly helpful. The views expressed here are those of the author and do not reflect the views of the U.S. Environmental Protection Agency.

### LITERATURE CITED

Allan, J. D. 1995. Stream ecology structure and function of running waters. Chapman and Hall, London, UK.

- Bennett, E. M., S. R. Carpenter, and N. F. Caraco. 2001. Human impact on erodable phosphorus and eutrophication: a global perspective. *BioScience* 51:227-234.
- Bergfur, J., R. K. Johnson, L. Sandin, W. Goedkoop, and K. Nygren. 2007. Effects of nutrient enrichment on boreal streams: invertebrates, fungi and leaf-litter breakdown. *Freshwater Biology* 52:1618-1633.
- Chételat, J., F. R. Pick, A. Morin, and P. B. Hamilton. 1999. Periphyton biomass and community composition in rivers of different nutrient status. *Canadian Journal of Fisheries and Aquatic Sciences* 56:560-569.
- Clenaghan, C., P. S. Giller, J. O'Halloran, and R. Hernan. 1998. Stream macroinvertebrate communities in a conifer-afforested catchment in Ireland: relationships to physico-chemical and biotic factors. *Freshwater Biology* 40:175-193.
- Correll, D. L. 1998. The role of phosphorus in the eutrophication of receiving waters: a review. *Journal of Environmental Quality* 27:261-266.
- Covich, A. P., M. A. Palmer, and T. A. Crowl. 1999. The role of benthic species in freshwater ecosystem processes. *BioScience* 49:119-126.
- Cross, W. F., B. R. Johnson, J. B. Wallace, and A. D. Rosemond. 2005. Contrasting response of stream detritivores to long-term nutrient enrichment. *Limnology and Oceanography* 50:1730-1739.
- Cross, W. F., J. B. Wallace, A. D. Rosemond, and S. L. Eggert. 2006. Whole-system nutrient enrichment increases secondary production in a detritus-based ecosystem. *Ecology* 87:1556-1565.
- Daly, C., G. H. Taylor, W. P. Gibson, T. W. Parzybok, G. L. Johnson, and P. Pasteris. 2001. High-quality spatial climate data sets for the United States and beyond. *Transactions of the American Society of Agricultural Engineers* 43:1957-1962.
- Dehejia, R. H., and S. Wahba. 2002. Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics* 84:151-161.
- Dodds, W. K., and E. B. Welch. 2000. Establishing nutrient criteria in streams. *Journal of the North American Benthological Society* 19:186-196.
- Dodson, S. I., S. E. Arnott, and K. L. Cottingham. 2000. The relationship in lake communities between primary productivity and species richness. *Ecology* 81:2662-2679.
- Elser, J. J., M. E. S. Bracken, E. E. Cleland, D. S. Gruner, W. S. Harpole, H. Hillebrand, J. T. Ngai, E. W. Seabloom, J. B. Shurin, and J. E. Smith. 2007. Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. *Ecology Letters* 10:1135-1142.
- Elwood, J. W., J. D. Newbold, A. F. Trimble, and R. W. Stark. 1981. The limiting role of phosphorus in a woodland stream ecosystem: effects of P enrichment on leaf decomposition and primary producers. *Ecology* 62:146-158.
- Evans-White, M. A., and G. A. Lamberti. 2006. Stoichiometry of consumer-driver nutrient recycling across nutrient regimes in streams. *Ecology Letters* 9:1186-1197.
- Feminella, J. W., and C. P. Hawkins. 1995. Interactions between stream herbivores and periphyton: a quantitative analysis of past experiments. *Journal of the North American Benthological Society* 14:465-509.
- Fore, L. S., J. R. Karr, and R. W. Wiseman. 1996. Assessing invertebrate responses to human activities: evaluating alternative approaches. *Journal of the North American Benthological Society* 15:212-231.
- Francoeur, S. N. 2001. Meta-analysis of lotic nutrient amendment experiments: detection and quantifying subtle responses. *Journal of the North American Benthological Society* 20:358-368.
- Frissell, C. A., W. J. Liss, C. E. Warren, and M. D. Hurley. 1986. A hierarchical framework for stream habitat classification: viewing streams in a watershed context. *Environmental Management* 10:199-214.
- Fuller, R. L., B. P. Kennedy, and C. Nielsen. 2004. Macroinvertebrate responses to algal and bacterial manipulations in streams. *Hydrobiologia* 523:113-126.
- Gafner, K., and C. T. Robinson. 2007. Nutrient enrichment influences the responses of stream macroinvertebrates to disturbance. *Journal of the North American Benthological Society* 26:92-102.
- Gulis, V., A. D. Rosemond, K. Suberkropp, H. S. Weyers, and J. P. Benstead. 2004. Effects of nutrient enrichment on the decomposition of wood and associated microbial activity in streams. *Freshwater Biology* 49:1437-1447.
- Gulis, V., and K. Suberkropp. 2004. Effects of whole-stream nutrient enrichment on the concentration and abundance of aquatic hyphomycete conidia in transport. *Mycologia* 96:57-65.
- Haddad, N. M., J. Haarstad, and D. Tilman. 2000. The effects of long-term nitrogen loading on grassland insect communities. *Oecologia* 124:73-84.
- Harding, J. S., R. G. Young, J. W. Hayes, K. A. Shearer, and J. D. Stark. 1999. Changes in agricultural intensity and river health along a river continuum. *Freshwater Biology* 42:345-357.
- Hart, D. D., and C. T. Robinson. 1990. Resource limitation in a stream community: phosphorus enrichment effects on periphyton and grazers. *Ecology* 71:1494-1502.
- Heino, J., T. Muotka, and R. Paavola. 2003. Determinants of macroinvertebrate diversity in headwater streams: regional and local influences. *Journal of Animal Ecology* 72:425-434.
- Hillebrand, H. 2002. Top-down versus bottom-up control of autotrophic biomass—a meta-analysis of experiments with periphyton. *Journal of the North American Benthological Society* 21:349-369.
- Holomuzki, J. R., and B. J. F. Biggs. 2006. Food limitation affects algalivory and grazer performance for New Zealand stream macroinvertebrates. *Hydrobiologia* 561:83-94.
- Imai, K., and D. A. Van Dyk. 2004. Causal inference with general treatment regimes: generalizing the propensity score. *Journal of the American Statistical Association* 99:854-866.
- Joffe, M. M., and P. R. Rosenbaum. 1999. Invited commentary: propensity scores. *American Journal of Epidemiology* 150:327-333.
- Kaplan, L. A., and T. L. Bott. 1989. Diel fluctuations in bacterial activity on streambed substrata during vernal algal blooms: effects of temperature, water chemistry, and habitat. *Limnology and Oceanography* 34:718-733.
- Kerans, B. L., and J. R. Karr. 1994. A benthic index of biotic integrity (B-IBI) for rivers of the Tennessee Valley. *Ecological Applications* 4:768-785.
- Lemly, A. D. 1998. Bacterial growth on stream insects: potential for use in bioassessment. *Journal of the North American Benthological Society* 17:228-238.
- Liess, A., and H. Hillebrand. 2006. Role of nutrient supply in grazer-periphyton interactions: reciprocal influences of periphyton and grazer nutrient stoichiometry. *Journal of the North American Benthological Society* 25:632-642.
- Maldonado, G., and S. Greenland. 2002. Estimating causal effects. *International Journal of Epidemiology* 31:422-429.
- McCutchan, J. H., and W. M. Lewis, Jr. 2002. Relative importance of carbon sources for macroinvertebrates in a Rocky Mountain stream. *Limnology and Oceanography* 47:742-752.
- Merritt, R. W., and K. W. Cummins. 1996. An introduction to the aquatic insects of North America. Third edition. Kendall Hunt, Dubuque, Iowa, USA.
- Mihuc, T. B., and G. W. Minshall. 1995. Trophic generalists vs. trophic specialists: implications for food web dynamics in post-fire streams. *Ecology* 76:2361-2372.

- Miltner, R. J., and E. T. Rankin. 1998. Primary nutrients and the biotic integrity of rivers and streams. *Freshwater Biology* 40:145–158.
- Morris, D. P., and W. M. Lewis, Jr. 1992. Nutrient limitation of bacterioplankton growth in Lake Dillon, Colorado. *Limnology and Oceanography* 37:1179–1192.
- Moss, D. M., M. T. Furse, J. F. Wright, and P. D. Armitage. 1987. The prediction of the macro-invertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. *Freshwater Biology* 17:41–52.
- Niyogi, D. K., M. Koren, C. J. Arbuckle, and C. R. Townsend. 2007. Stream communities along a catchment land-use gradient: subsidy-stress responses to pastoral development. *Freshwater Biology* 39:213–225.
- Peck, D. V., J. M. Lazorchak, and D. J. Klemm. 2008. Environmental monitoring and assessment- program surface waters: western pilot study field operations manual for Wadeable streams. U.S. Environmental Protection Agency, Washington, D.C., USA. (<http://www.epa.gov/emap/html/pubs/docs/groupdocs/surfwater/field/ewwsm01.html>)
- Quinn, J. M., A. B. Cooper, M. J. Stroud, and G. P. Burrell. 1997. Shade effects on stream periphyton and invertebrates: an experiment in streamside channels. *New Zealand Journal of Marine and Freshwater Research* 31:665–683.
- R Development Core Team. 2008. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (<http://www.R-project.org>)
- Richards, C., L. B. Johnson, and G. E. Host. 1996. Landscape-scale influences on stream habitats and biota. *Canadian Journal of Fisheries and Aquatic Science* 53(Supplement 1): 295–311.
- Rosemond, A. D., P. J. Mulholland, and J. W. Elwood. 1993. Top-down and bottom-up control of stream periphyton: effects of nutrients and herbivores. *Ecology* 74:1264–1280.
- Rosenbaum, P. R. 2002. *Observational studies*. Second edition. Springer, New York, New York, USA.
- Rosenbaum, P. R., and D. B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70:41–55.
- Roy, A. H., A. D. Rosemond, M. J. Paul, D. S. Leigh, and J. B. Wallace. 2003. Stream macroinvertebrate response to catchment urbanization (Georgia, USA). *Freshwater Biology* 48: 329–346.
- Sabater, S., V. Acuna, A. Giorgi, E. Guerra, I. Munoz, and A. M. Romani. 2005. Effects of nutrient inputs in a forested Mediterranean stream, under moderate light availability. *Archiv für Hydrobiologie* 163:479–496.
- Shipley, B. 1999. Testing causal explanations in organismal biology: causation, correlation and structural equation modeling. *Oikos* 86:374–382.
- Slavik, K., B. J. Peterson, L. A. Deegan, W. B. Bowden, A. E. Hershey, and J. E. Hobbie. 2004. Long-term responses of the Kuparuk River ecosystem to phosphorus fertilization. *Ecology* 85:939–954.
- Smil, V. 2000. Phosphorus in the environment: natural flows and human interferences. *Annual Review of Energy and the Environment* 25:53–88.
- Smith, H. L. 1997. Matching with multiple controls to estimate treatment effects in observational studies. *Sociological Methodology* 27:325–353.
- Smith, V. H. 1979. Nutrient dependence of primary productivity in lakes. *Limnology and Oceanography* 24:1051–1064.
- Sterner, W. R., and J. J. Elser. 2002. *Ecological stoichiometry*. Princeton University Press, Princeton, New Jersey and Oxford, UK.
- Stoddard, J. L., et al. 2006. Environmental monitoring and assessment (EMAP) western streams and rivers statistical summary. EPA/620/R-05/006. Office of Research and Development, U.S. Environmental Protection Agency, Washington, D.C., USA.
- Tank, J. L., and W. K. Dodds. 2003. Nutrient limitation of epilithic and epixylic biofilms in ten North American streams. *Freshwater Biology* 48:1031–1049.
- Tilman, D. 1987. Secondary succession and the pattern of plant dominance along experimental nitrogen gradients. *Ecological Monographs* 57:189–214.
- U.S. EPA (U.S. Environmental Protection Agency). 2004. Wadeable streams assessment water chemistry laboratory manual. EPA841-B-04-008. Office of Water, Office of Environmental Information, Washington, D.C., USA.
- Valiela, I., K. Foreman, M. LaMontagne, D. Hersh, J. Costa, P. Peckol, B. DeMeo-Andreson, C. D'Avanzo, M. Babione, C. Sham, J. Brawley, and K. Lajtha. 1992. Couplings of watersheds and coastal waters: sources and consequences of nutrient enrichment in Waquoit Bay, Massachusetts. *Estuaries* 15:443–457.
- Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing. 1980. The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences* 37:130–137.
- Vieira, N. K. M., N. L. Poff, D. M. Carlisle, S. R. Moulton II, M. L. Koski, and B. C. Kondratieff. 2006. A database of lotic invertebrate traits for North America. U.S. Geological Survey Data Series 187. (<http://pubs.water.usgs.gov/ds187>)
- Vinson, M. R., and C. P. Hawkins. 2003. Broad-scale geographical patterns in local stream insect genera richness. *Ecography* 26:751–767.
- Vitousek, P. M., J. D. Aber, R. W. Howarth, G. E. Likens, P. A. Matson, D. W. Schindler, W. H. Schlesinger, and D. G. Tilman. 1997. Human alteration of the global nitrogen cycle: sources and consequences. *Ecological Applications* 7:737–750.
- Vogelmann, J. E., S. M. Howard, L. Yang, C. R. Larson, B. K. Wylie, and N. Van Driel. 2001. Completion of the 1990s national land cover data set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. *Photogrammetric Engineering and Remote Sensing* 67:650–652.
- Wang, L., D. M. Robertson, and P. J. Garrison. 2007. Linkages between nutrients and assemblages of macroinvertebrates and fish in Wadeable streams. *Environmental Management* 39:194–212.
- Wellnitz, T. A., R. B. Rader, and J. V. Ward. 1996. Light and a grazing mayfly shape periphyton in a Rocky Mountain stream. *Journal of the North American Benthological Society* 15:496–507.
- Woo, M.-J., J. P. Reiter, and A. F. Karr. 2008. Estimation of propensity scores using generalized additive models. *Statistics in Medicine* 27:3805–3816.
- Wood, S. N., and N. H. Augustin. 2002. GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological Modelling* 157:157–177.
- Yuan, L. L., and S. B. Norton. 2003. Comparing responses of macroinvertebrate metrics to increasing stress. *Journal of the North American Benthological Society* 22:308–322.

## APPENDIX

Table of weightings applied to different types of geographical information to compute the grazing intensity index (*Ecological Archives* A020-002-A1).